ON THE DETERMINATION OF THE NUMBER OF SIGNALS AND ITS PERFORMANCE ANALYSIS IN PRESENCE OF WHITE NOISE

Debasis Kundu

Technical Report 99-12

July 1999

Center for Multivariate Analysis
417 Thomas Building
Penn State University
University Park, PA 16802

Research work of authors was partially supported by the Army Research Office under Grant DAAHO4-96-1-0082. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 21, 1999	Technical - July 1999

4. TITLE AND SUBTITLE
On the Determination of the Number of Signals and its Performance Analysis in Presence of White Noise

5. FUNDING NUMBERS
DAAH04-96-1-0082

6. AUTHOR(S)
D. Kundu

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Center for Multivariate Analysis
417 Thomas Building
Department of Statistics
Penn State University
University Park, PA 16802

8. PERFORMING ORGANIZATION REPORT NUMBER
99-12

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 35518.57-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* In signal processing, high resolution signal parameter estimation is a significant problem. In particular the estimation of the direction of the narrow band signals emitted by multiple sources received wide applications recently in signal processing literature. Quite a number of papers appeared in the lasy twenty five years regarding the estimation of the parameters of the direction of arrival of signals, but not that much attention has been given in estimating the number of signals. In this paper we develop a method using penalty function technique. But instead of using any fixed penalty function like AIC or MDL, a class of penalty functions sarusfying some special properties have been used. We prove that any penalty function from the particular class will produce consistent estimates under the assumptions that the error random variables are independent and identical distributed with mean zero and finite variance. We also obtain the probabilities of wrong detection for any particular penalty function and estimate it using the matrix perturbation technique. It gives some idea to choose the proper penalty function for any particular model. Simulations are performed to verify the usefulness of the analysis and to compare our method with the existing ones.

14. SUBJECT TERMS Almost sure convergence, Information theoretic criteria, model selection, signal detection

15. NUMBER IF PAGES
9

16. PRICE CODE

17 SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ON THE DETERMINATION OF THE NUMBER OF SIGNALS AND ITS PERFORMANCE ANALYSIS IN PRESENCE OF WHITE NOISE

Debasis Kundu
Department of Mathematics
Indian Institute of Technology Kanpur
Kanpur, Pin 208016, India
kundu@iitk.ac.in

Abstract: In signal processing, high resolution signal parameter estimation is a significant problem. In particular the estimation of the direction of the narrow band signals emitted by multiple sources received wide applications recently in signal processing literature. Quite a number of papers appeared in the last twenty five years regarding the estimation of the parameters of the direction of arrival of signals, but not that much attention has been given in estimating the number of signals. In this paper we develop a method using penalty function technique. But instead of using any fixed penalty function like AIC or MDL, a class of penalty functions satisfying some special properties have been used. We prove that any penalty function from that particular class will produce consistent estimates under the assumptions that the error random variables are independent and identically distributed with mean zero and finite variance. We also obtain the probabilities of wrong detection for any particular penalty function and estimate it using the matrix perturbation technique. It gives some idea to choose the proper penalty function for any particular model. Simulations are performed to verify the usefulness of the analysis and to compare our method with the existing ones.

AMS Subject Classifications: 62H12, 62H15, 62F12

Keywords and Phrases: Almost sure convergence, Information theoretic criteria, model selection, signal detection.

Short Running Title: Detecting the number of signals.

e-mail: kundu@stat.psu.edu.

1. INTRODUCTION:

Detecting the number of signals and estimating the parameters of the signals are important problems in signal processing. There has been a great deal of recent interests in the use of signal subspace processing methods for the estimation of the direction of arrival (DOA) of multiple plane waves or frequencies of sinusoids. See for example, [1, 2, 7, 10, 11, 12, 14, 16, 23, 25, 28]. The problem can be formulated as follows:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t); \qquad t = 1, \ldots, N, \qquad (1.1)$$

where $\mathbf{x}(t)$ is a $p \times 1$ complex valued observation vector, $\mathbf{s}(t)$ is the $q \times 1$ complex valued unobserved signal vector and $\mathbf{n}(t)$ is a $p \times 1$ complex valued noise vector at the time point t. $\mathbf{A} = [\, \mathbf{A}(\phi_1) \ldots \mathbf{A}(\phi_q)]$ is a $p \times q$ matrix, where $\mathbf{A}(\phi_k)$ is the $p \times 1$ complex valued direction vector of the k^{th} wavefront and parameterized by an unknown parameter vector ϕ_k, associated with the kth signal. We assume that $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are complex vectors which are distributed independently with each other. The $p \times q$ matrix $\mathbf{A}$ has the special structure

$$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_q]$$

and

$$\mathbf{a}_k = \mathbf{a}(\phi_k) = [1, e^{-j\omega_0\phi_k}, \ldots, e^{-j\omega_0(p-1)\phi_k}]^T, \qquad (1.2)$$

where $j = \sqrt{-1}$, $\phi_k = C^{-1}\Lambda sin(\theta_k)$, C = speed of propagation, θ_k is the direction of arrival of signal from the k^{th} source and Λ is the inter-source distance (see [17]). It is always possible to take ω_0 to be unity, without loss of generality (see [17]). One of the important problem is the estimation of q, the number of signals another is the estimation of $\phi_1, \ldots, \phi_q$.

Estimating the parameters of the model (1.1) is a very important problem in signal processing. This is the situation in sensor array processing [5, 6, 10], in Harmonic analysis [15], in retrieving the poles of a system from natural response (Wax, Schmidt and Kailath; [24]) and also in retrieving overlapping echoes from radar back scatter ([17]).

Estimation of $\phi_1, \ldots, \phi_q$ assuming q known is usually solved by some eigen decomposition method. There are several eigen-decomposition methods available in the literature, for example MUSIC [3, 22], modified MUSIC [12], ESPRIT, TLS-ESPRIT ([20]), GEESE

([17]), Bai and Rao method [1] , the modified Bai and Rao method [14] or the method proposed by Kannan, Kundu and Mitra [8]. For detailed discussions of the different eigen-decomposition methods the readers are referred to the Ph.D. thesis of Kannan [7] or the review article of Paulraj *et al.* [16]. MUSIC and modified MUSIC algorithms are obtained by minimizing the Hermitian form of an exponential function. The solution is obtained by a search procedure which is iterative in nature. The other decomposition methods like ESPRIT, TLS-ESPRIT, GEESE, Kannan, Kundu and Mitra method, Bai and Rao method or the modified Bai and Rao method are all non iterative in nature. It is observed that among several non iterative methods centro symmetric modified Bai and Rao method works very well.

The estimation of q was attempted by many researchers. Wax and Kailath [23] and Zhao, Krishnaiah and Bai [28] study this problem from the parametric point of view, where as Bhandari and Bansal [2] and Yin and Krishnaiah [27] study this problem from the Bayesian and non parametric point of view respectively. Comparison and comments on the different methods can be found in [11].

Estimation of the number of signals and the performance analysis of a related model can be found in [19, 22, 9, 13]. In all the methods, *i.e.* in estimation and in performance analysis computation, it is assumed that the signal random variable $\mathbf{s}(t)$ and the noise random variable $\mathbf{n}(t)$ are Gaussian random variables and it is not very easy to relax this assumption.

The main aim of this paper is to provide a consistent method of estimation of q of the model (1.1) and carry out the performance analysis without assuming that $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are Gaussian random variables. We assume the following

$$E(\mathbf{s}(t)) = \mathbf{0}, \qquad E(\mathbf{s}(t)\mathbf{s}(t)^H) = \Psi > 0, \quad (1.3)$$

$$E(\mathbf{n}(t)) = \mathbf{0}, \qquad E(\mathbf{n}(t)\mathbf{n}(t)^H) = \sigma^2\mathbf{I},$$

$\mathbf{s}(t)$ and $\mathbf{n}(t)$ are independent and $(\mathbf{s}(t), \mathbf{n}(t))$; for $t = 1, \ldots, N$ are independent and identically distributed random variables. Here $\mathbf{I}$ is the identity matrix of order $p \times p$ and 'H' denotes the conjugate transpose of a matrix or a vector. In developing the procedure, we use the information theoretic criteria. But not any fixed penalty function has been used like AIC or MDL, but a class of penalty functions satisfying some special properties like EDC of [28] has been used. Unlike [28] proving the strong consistency we don't need any distributional assumption of $\mathbf{s}(t)$ or $\mathbf{n}(t)$. We carry out the performance analysis of the proposed method using the matrix perturbation technique and large sample approximation. We compute the probability of wrong detection for large sample size for any particular penalty function in that given class and that gives some idea which penalty function should be used for any particular sample.

The organization of the rest of the paper is as follows. We develop the method in Section 2 and the strong consistency results are provided in Section 3. The performance analysis is carried out in Section 4 and some numerical results are reported in Section 5. The choice of the penalty function is suggested in Section 6. In Section 7 we address the problem if the error is known to be Gaussian and finally we draw conclusions from our work in Section 8.

2. SOME NOTATIONS AND ESTIMATION PROCEDURE:

We use the following notations through out the paper. Let $\mathbf{R}$ be the variance covariance matrix, *i.e.*

$$\mathbf{R} = E(\mathbf{x}(t)\mathbf{x}(t)^H) = \mathbf{A}\Psi\mathbf{A}^H + \sigma^2\mathbf{I} \quad (2.1)$$

and $\hat{\mathbf{R}}$ be the sample variance covariance matrix, *i.e.*

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}(t)^H. \quad (2.2)$$

Although $\hat{\mathbf{R}}$ depends on N, for brevity we are not making it explicit. We can write

$$\hat{\mathbf{R}} = \mathbf{R} + (\hat{\mathbf{R}} - \mathbf{R}). \quad (2.3)$$

Here $(\hat{\mathbf{R}} - \mathbf{R})$ denotes the perturbation of the matrix $\mathbf{R}$ and clearly some norm of the perturbation matrix goes to zero as N tends to infinity. Let's denote the spectral decomposition of the matrix $\mathbf{R}$ be as follows

$$\mathbf{R} = \sum_{i=1}^{p}\lambda_{(i)}\mathbf{Z}_i\mathbf{Z}_i^H, \quad \lambda_{(1)} > \ldots > \lambda_{(q)} > \lambda_{(q+1)}$$

$$= \ldots = \lambda_{(p)} = \sigma^2. \quad (2.4)$$

Here $\lambda_{(i)}$'s are the eigenvalues and $\mathbf{Z}_i$'s are corresponding orthonormal eigenvectors of $\mathbf{R}$. Note that $\lambda_{(1)}, \ldots, \lambda_{(p)}$ denote the ordered eigenvalues of $\mathbf{R}$. Let the corresponding spectral decomposition of $\mathbf{R}$ be as follows:

$$\hat{\mathbf{R}} = \sum_{i=1}^{p}\hat{\lambda}_{(i)}\hat{\mathbf{Z}}_i\hat{\mathbf{Z}}_i^H, \qquad \hat{\lambda}_{(1)} > \ldots > \hat{\lambda}_{(p)}, \quad (2.5)$$

here $\hat{\lambda}_{(i)}$'s are the eigen values and $\hat{\mathbf{Z}}_i$'s are the corresponding orthonormal eigen vectors of $\hat{\mathbf{R}}$. Similarly $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(p)}$ are the ordered eigen values of $\hat{\mathbf{R}}$. Consider the following function

$$IC(k, C_N) = \hat{\lambda}_{(k+1)} + kC_N; \qquad k = 0, 1, \ldots p-1, \quad (2.6)$$

here, C_N satisfies the following conditions

$$(a)\ C_N > 0 \qquad (b)\ C_N \to 0 \quad (c)\ \frac{\sqrt{N}C_N}{\sqrt{loglogN}} \to \infty. \tag{2.7}$$

Let

$$\hat{q} = \arg\min\ IC(k, C_N), \qquad \text{for} \quad 0 \le k \le p-1, \tag{2.8}$$

then $\hat{q}$ is an estimator of q. Clearly $\hat{q}$ is a function of N and C_N, but we are not making it explicit for notational convenience. In the next section we prove that $\hat{q}$ is a consistent estimator of q if $(\mathbf{s}(t), \mathbf{n}(t))$ satisfies assumption (1.3) and C_N satisfies (2.7). In the subsequent section we suggest how to choose C_N.

3. CONSISTENCY RESULTS:

We need the following lemma for further development.

Lemma 1: Let $\mathbf{P} = ((P_{ij}))$ and $\mathbf{Q} = ((Q_{ij}))$ be two $m \times m$ Hermitian matrices with the following spectral decompositions

$$\mathbf{P} = \sum_{i=1}^{m} \delta_i \mathbf{u}_i \mathbf{u}_i^H, \qquad \mathbf{Q} = \sum_{i=1}^{m} \mu_i \mathbf{v}_i \mathbf{v}_i^H, \tag{3.1}$$

where $\delta_1 \ge \ldots \ge \delta_m$ and $\mu_1 \ge \ldots \ge \mu_m$
and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1 \ldots \mathbf{v}_m\}$ are the orthonormal set of eigenvectors of $\mathbf{P}$ and $\mathbf{Q}$ respectively. If there exists an α, such that $|P_{ij} - Q_{ij}| \le \alpha$ for all $i, j = 1, \ldots, m$, then there exists a C such that $|\delta_i - \mu_i| \le C\alpha$ for all $i = 1, \ldots m$.

Proof: The proof mainly follows from von Neumann inequality but see also [1] for details.

By the law of iterated logarithm, we can say that

$$\hat{\mathbf{R}} = \mathbf{R} + O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}} \quad a.s., \tag{3.2}$$

If we denote the non zero eigenvalues of $\mathbf{A}\mathbf{\Psi}\mathbf{A}^H$ as $\gamma_1, \ldots, \gamma_q$ (order them as $\gamma_{(1)} > \ldots > \gamma_{(q)}$) then it is immediate that

$$\begin{aligned} \lambda_{(i)} &= \gamma_{(i)} + \sigma^2 \quad \text{for} \quad i = 1, \ldots q \\ \lambda_{(i)} &= \sigma^2 \quad \text{for} \quad i = q+1, \ldots, p. \end{aligned} \tag{3.3}$$

Observe that to prove $\hat{q}$ is a strongly consistent estimator of q, it is enough to prove that for large N,

$$IC(q, C_N) - IC(k, C_N) < 0 \quad a.s. \tag{3.4}$$

for $k = 0, 1, \ldots, q-1, q+1, \ldots, p-1$. Consider two different cases

Case 1: $k < q$,

$$\begin{aligned} & IC(q, C_N) - IC(k, C_N) \\ &= \hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} + (q-k)C_N \\ &= \lambda_{(q+1)} - \lambda_{(k+1)} + (q-k)C_N + O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}} \\ &= -\gamma_{(k+1)} + (q-k)C_N + O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}}. \end{aligned} \tag{3.5}$$

Since the second and the third term of (3.5) go to zero as N tends to infinity and since $\gamma(k+1) > 0$, implies for large N,

$$IC(q, C_N) - IC(k, C_N) < 0 \quad a.s.. \tag{3.6}$$

case II $q < k$ $IC(q, C_N) - IC(k, C_N)$
$= \hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} + (q-k)C_N$
$= (q-k)C_N + O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}}$.
Therefore

$$\frac{(IC(q, C_N) - IC(k, C_N)}{C_N}$$

$$= (q-k) + \frac{1}{C_N} O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}}. \tag{3.7}$$

Now by the properties of C_N, the second term on the right hand side goes to zero. Since $C_N > 0$, therefore for large N

$$IC(q, C_N) - IC(k, C_N) < 0 \qquad a.s. \tag{3.8}$$

for $q < k$. Combining (3.6) and (3.8), we obtain (3.4) and that proves the result.

4. PERFORMANCE ANALYSIS

In this section we obtain the bound for $P\{\hat{q} \ne q\}$ at least for large N. Note that

$$P\{\hat{q} \ne q\} = \sum_{k=0}^{q-1} P\{\hat{q} = k\} + \sum_{k=q+1}^{p-1} P\{\hat{q} = k\}. \tag{4.1}$$

Consider two different cases:
Case I If $k < q$

$$\begin{aligned} & P\{\hat{q} = k\} \\ &= P\{IC(k, C_N) < IC(j, C_N) \\ & \quad \text{for } j = 0, 1, \ldots, k-1, k+1, \ldots, p-1\} \end{aligned}$$

$$
\begin{aligned}
&\leq P\{IC(k,C_N) < IC(q,C_N)\} \\
&= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(q+1)} + (k-q)C_N < 0\} \\
&= P\{\lambda_{(k+1)} - \lambda_{(q+1)} + (k-q)C_N \\
&< (\lambda_{(k+1)} - \hat{\lambda}_{(k+1)} + \hat{\lambda}_{(q+1)} - \lambda_{(q+1)})\} \\
&= P\{\gamma_{(k+1)} \\
&< (q-k)C_N + (\lambda_{(k+1)} - \hat{\lambda}_{(k+1)}) + \\
&\qquad \hat{\lambda}_{(q+1)} - \lambda_{(q+1)}\}.
\end{aligned} \tag{4.2}
$$

Since C_N tends to zero, $(\lambda_{(k+1)} - \hat{\lambda}_{(k+1)})$ tends to zero, $(\hat{\lambda}_{(q+1)} - \lambda_{(q+1)})$ tends to zero and $\gamma_{(k+1)} > 0$, *a.s.*, therefore for sufficiently large N,

$$
P\{\hat{q} = k\} = 0 \qquad \text{for} \qquad k < q. \tag{4.3}
$$

Case II If $k > q$

$$
\begin{aligned}
P\{\hat{q} = k\} &= P\{IC(k,C_N) < IC(j,C_N) \\
&= \text{for } j = 0,\ldots,k-1,k+1,\ldots,p-1\} \\
&= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k-j)C_N < 0 \\
&\quad \text{for } j = 0,1,\ldots,k-1,k+1,\ldots,p-1\}
\end{aligned}
$$

Therefore, by the same argument as (4.3), we can say that for large N,

$$
\begin{aligned}
P(\hat{q} = k) &= P\{\lambda_{(j+1)} - \lambda_{(k+1)} + O\left(\frac{loglogN}{N}\right)^{\frac{1}{2}} \\
&> (k-j)C_N \\
&\quad \text{for } j = 0,1,\ldots,k-1,k+1,\ldots,p-1\} \\
&= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k-j)C_N < 0 \\
&\quad \text{for } j = q,\ldots,k-1,k+1,\ldots,p-1\} \\
&\leq P\{\hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} > (k-q)C_N\}
\end{aligned}
$$

since $\lambda_{(j+1)} - \lambda_{(k+1)} > 0$ for $j = 0,1,\ldots,q-1$. So, for large N

$$
P\{\hat{q} \neq q\} \approx \sum_{k=q+1}^{p-1} P(\hat{q} = k). \tag{4.4}
$$

Therefore, to compute (4.4), we need to know the joint distribution of $\hat{\lambda}_{(1)},\ldots,\hat{\lambda}_{(p)}$. We use matrix perturbation technique to compute the joint distribution of $\hat{\lambda}_{(1)},\ldots,\hat{\lambda}_{(p)}$. Observe that from the Central limit theorem, we can say;

$$
\sqrt{N}(Vec(\hat{\mathbf{R}}) - Vec(\mathbf{R}))
$$

is asymptotically normal with mean vector $\mathbf{0}$ and certain $p^2 \times p^2$ dispersion matrix $\mathbf{\Gamma}$. Here $Vec(.)$ of a $p \times p$ matrix is a $p^2 \times 1$ vector obtained by stacking one column below another. Let's write

$$
\hat{\mathbf{R}} = \mathbf{R} + (\hat{\mathbf{R}} - \mathbf{R}) = \mathbf{R} + \epsilon_N \frac{\hat{\mathbf{R}} - \mathbf{R}}{\epsilon_N} = \mathbf{R} + \epsilon_N \mathbf{B}_N, \tag{4.5}
$$

here $\epsilon_N = \{loglogN/N\}^{\frac{1}{2}}$, therefore for large N, $0 < \epsilon_N < 1$ and the elements of $\mathbf{B}_N$'s are bounded almost surely because of (3.2). Let λ_i be any particular eigenvalue of $\mathbf{R}$ and $\hat{\lambda}_i$ be the corresponding perturbed eigenvalue of $\hat{\mathbf{R}}$. Suppose $\mathbf{Z}_i$ is the normalized eigenvector of $\mathbf{R}$ corresponding to λ_i, then from [26], we have

$$
\hat{\lambda}_i \approx \lambda_i + \epsilon_N \mathbf{Z}_i^H \mathbf{B}_N \mathbf{Z}_i. \tag{4.6}
$$

It is important to note that λ_i may be repeated eigenvalue, then $\mathbf{Z}_i$ is not unique. Take any particular $\mathbf{Z}_i$, still (4.6) is valid ([26]). Since the elements of $\mathbf{B}_N$ are asymptotically normally distributed, therefore $\mathbf{Z}_i^H \mathbf{B}_N \mathbf{Z}_i$ will also be asymptotically normally distributed. Clearly $E(\hat{\lambda}_i) = \lambda_i$ for $i = 1,\ldots,p$ as $E(\mathbf{B}_N) = 0$ and

$$
E(\hat{\lambda}_i - \lambda_i)(\hat{\lambda}_j - \lambda_j) = \epsilon_N^2 E(\mathbf{Z}_i^H \mathbf{B}_N \mathbf{Z}_i)(\mathbf{Z}_j^H \mathbf{B}_N \mathbf{Z}_j), \tag{4.7}
$$

where $\mathbf{Z}_i$ and $\mathbf{Z}_j$ are two orthonormal eigenvectors corresponding to λ_i and λ_j. Now

$$
\begin{aligned}
&E(\mathbf{Z}_i^H \mathbf{B}_N \mathbf{Z}_i)(\mathbf{Z}_j^H \mathbf{B}_N \mathbf{Z}_j) \\
&= \frac{1}{\epsilon_N^2} E[\mathbf{Z}_i^H (\hat{\mathbf{R}}_N - \mathbf{R}) \mathbf{Z}_i \mathbf{Z}_j^H (\hat{\mathbf{R}}_N - \mathbf{R}) \mathbf{Z}_j] \\
&= \frac{1}{N\epsilon_N^2} (\mathbf{Z}_i^H \mathbf{R} \mathbf{Z}_j)(\mathbf{Z}_j^H \mathbf{R} \mathbf{Z}_i).
\end{aligned} \tag{4.8}
$$

Note that the last equality of (4.8) follows from [4]. Therefore, from (4.8) it is clear that $\hat{\lambda}_i$ will be asymptotically normally distributed with mean λ_i and variance $\frac{\lambda_i^2}{N}$ for $i = 1,\ldots,p$. Asymptotically $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are independently distributed for $i \neq j$. Note that for large N, the distribution of $\{\hat{\lambda}_{(q+1)},\ldots,\hat{\lambda}_{(p)}\}$ is the distribution of order statistics of $(p-q)$ random random sample from a normal distribution with mean σ^2 and variance $\frac{\sigma^4}{N}$. Therefore, to compute the right hand side of (4.4) we need to know the joint distribution of $\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)}\}$ for $j = q\ldots,p-1$.

It is well known that the exact distribution of the difference of order statistics of normal distribution are difficult to obtain. Although some well known approximations results are available. We use re-sampling technique similarly as [13] to estimate (4.4). The details will be presented in the next section.

5. NUMERICAL EXPERIMENTS:

In this section we perform some numerical experiments to present both the effectiveness of our method and the usefulness of the analysis. All the computations are performed at the Pennsylvania State University using SUN workstation. We use the RAN2 uniform random number generator of [18] and the singular value decomposition by IMSL subroutine. The programs are

written in FORTRAN. It is available on request from the author. We consider the following model.

$$p = 5, q = 2, \phi_1 = 1.0, \phi_2 = 2.0$$

The covariance matrix of the real and imaginary part of $\mathbf{x(t)}$ is a 2×2 matrix as follows.

$$\begin{bmatrix} 1.25 & 1.00 \\ 1.00 & 1.25 \end{bmatrix}$$

We consider $N = 100$. The real and imaginary part of $\mathbf{x(t)}$ are taken to be independent. For the comparison purposes with the other known methods, we consider the error random variables to be normally distributed with $\sigma = 0.75$ (SNR $\approx$ 3.01dB), $\sigma = 1.0$ (SNR $\approx$ .511dB) and $\sigma = 1.125$ (SNR $\approx$ -.511dB).

We use twelve different C_N, all of them satisfying (2.7), but converging to zero at different rates. We define them as $P(1), \ldots, P(12)$, they are as follows $P(1) = (\frac{1}{N})^{.1}$, $P(2) = (\frac{1}{N})^{.2}$, $P(3) = (\frac{1}{N})^{.3}$,
$P(4) = (\frac{1}{N})^{.4}$, $P(5) = \frac{1}{logN}$, $P(6) = \left(\frac{1}{logN}\right)^{.2}$,
$P(7) = \left(\frac{1}{logN}\right)^{.4}$, $P(8) = \left(\frac{1}{logN}\right)^{.6}$, $P(9) = \left(\frac{1}{logN}\right)^{.9}$,
$P(10) = \left(\frac{1}{NlogN}\right)^{.1}$, $P(11) = \left(\frac{1}{NlogN}\right)^{.3}$,
$P(12) = \left(\frac{1}{loglogN}\right)$.
Out of 1000 replications, the percentage of correct estimates (PCE), and the percentage of wrong estimates (PWE) are obtained for different SNR. We also obtain the theoretical values of the probabilities as follows. We draw a sample of size $(p - q)$ from a Gaussian random variable with mean σ^2 and variance $\frac{\sigma^4}{N}$. We order them as $\hat{\lambda}_{(q+1)}, \ldots, \hat{\lambda}_{(p)}$ and check whether

$$\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k - j)C_N < 0$$

for $j = q, \ldots, k-1, k+1, \ldots, p$. We repeat the process five thousand times and compute the percentage of time it is true and that gives an estimate of (4.4). We also calculate the percentage of under estimates from the distributional properties of $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(q)}$. Adding the probability of over estimate and the probability of under estimate we obtain the probability of wrong estimate. Finally subtracting the probability of wrong estimate from one we obtain the probability of correct estimate. The results are reported in Table 1. The quantity within the bracket indicates the theoretical estimate of the probability of correct estimate (PCE) and the probability of wrong estimate (PWE).

At finite sample size the performance of the proposed method very much depends on the penalty function used, although all of them give consistent estimates as the sample size tends to infinity. From Table 1 it is clear that the performances of all the methods becomes worse at low SNR which is not very surprising. It is important to observe that the theoretical probabilities match quite well in almost all the cases considered and the estimates are better in most of the cases at high SNR.

Table 1

σ = .75

P(k)	PCE	PWE
P(1)	1.00 (1.00)	0.00 (0.00)
P(2)	0.95 (1.00)	0.05 (0.00)
P(3)	0.56 (0.74)	0.44 (0.26)
P(4)	0.12 (0.23)	0.88 (0.77)
P(5)	0.38 (0.56)	0.62 (0.44)
P(6)	1.00 (1.00)	0.00 (0.00)
P(7)	0.99 (1.00)	0.01 (0.00)
P(8)	0.96 (1.00)	0.04 (0.00)
P(9)	0.77 (0.89)	0.23 (0.11)
P(10)	1.00 (1.00)	0.00 (0.00)
P(11)	0.12 (0.23)	0.88 (0.77)
P(12)	1.00 (1.00)	0.00 (0.00)

σ = 1.00

PCE	PWE	
P(1)	0.92 (0.98)	0.08 (0.02)
P(2)	0.42 (0.61)	0.58 (0.39)
P(3)	0.07 (0.21)	0.93 (0.79)
P(4)	0.00 (0.11)	1.00 (0.89)
P(5)	0.03 (0.19)	0.97 (0.81)
P(6)	0.97 (1.00)	0.03 (0.00)
P(7)	0.81 (0.96)	0.19 (0.04)
P(8)	0.42 (0.60)	0.58 (0.40)
P(9)	0.15 (0.29)	0.85 (0.71)
P(10)	0.80 (0.96)	0.20 (0.04)
P(11)	0.00 (0.14)	1.00 (0.86)
P(12)	0.93 (0.99)	0.07 (0.01)

σ = 1.125

PCE	PWE	
P(1)	0.72 (0.88)	0.28 (0.12)
P(2)	0.20 (0.43)	0.80 (0.57)
P(3)	0.01 (0.21)	0.99 (0.79)
P(4)	0.00 (0.19)	1.00 (0.81)
P(5)	0.00 (0.22)	1.00 (0.78)
P(6)	0.86 (0.96)	0.14 (0.04)
P(7)	0.53 (0.71)	0.47 (0.29)
P(8)	0.20 (0.41)	0.80 (0.59)
P(9)	0.04 (0.25)	0.96 (0.75)
P(10)	0.52 (0.70)	0.48 (0.30)
P(11)	0.00 (0.23)	1.00 (0.77)
P(12)	0.75 (0.91)	0.25 (0.09)

6. HOW TO CHOOSE THE PENALTY FUNCTION ?

Looking at the tables, it is clear that the theoretical bounds are quite close to the actual one. But unfortunately without knowing the actual parameters, we can't calculate the theoretical probabilities. Nobody, so far did raise this question that how to estimate these bounds. We estimate these probabilities with the help of the given sample and using re-sampling technique. We finally use them to choose the proper penalty function, which definitely depends on the model as well as the given sample. From any particular realization of the model, we compute the matrix $\hat{\mathbf{R}}$ (see (2.2)) and obtain the p eigenvalues and the corresponding eigenvectors. Now suppose using the penalty function $P(k)$, we estimate the order of the model as M_k. Assuming M_k is the correct order model, we compute the estimate of σ^2, by averaging the last $p - M_k$ eigenvalues, say $\hat{\sigma}^2$. We estimate (4.4), the probability of over estimate, assuming $\hat{\sigma}^2$ is the true value of σ^2, M_k is the correct order model and using the re-sampling (simulation) technique as described in the previous section by generating B normal random sample of size $p - M_k$. Similarly assuming $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(M_k)}$ are the true values of $\lambda_{(1)}, \ldots, \lambda_{(M_k)}$, M_k is the correct order model and using the asymptotic distribution of $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(M_k)}$, as obtained in section 4, we estimate the probability of under estimate. Therefore, adding the two we obtain an estimate of the probability of wrong detection. It can be shown easily that for large N, the estimate of probability of wrong detection under the assumptions of correct order model will be less than the estimate of wrong detection under the assumption of lower/ higher order model, because the former one goes to zero as N tends to infinity where as the later one goes to a positive quantity.

We use this idea and compute the estimate of the probability of wrong detection for all the criteria and choose that one which gives the lowest estimate of probability of wrong detection. We use the same model, and the same set of penalty functions and in each trial we choose that penalty function that gives the lowest estimate of probability of wrong detection. In each case we draw one hundred random sample ($B = 100$) to compute the probability of error and replicate it over one thousand trials. The result is reported in Table 2, which indicates the percentage of correct estimate (PCE) and the percentage of wrong estimate (PWE) over one thousand replications.

Table 2

σ	PCE	PWE
0.75	1.000	0.000
1.00	0.920	0.080
1.125	0.770	0.230

It is observed from Table 2 that the proposed method works quite well. As the SNR increases the performance of the proposed method improves, it verifies the consistency property of the proposed method. If the SNR is high ($\approx$ 3dB) then the proposed method can detect all the times the correct order model. As the SNR decreases the performance becomes bad. Even then at low SNR ($\approx$ -.5dB) the proposed method can detect more than 75 % of the time the correct order model.

7. IF THE ERROR IS GAUSSIAN:

So far we did not use any distributional assumptions on the error random variables of the model except (1.3). Zhao, Krishnaiah and Bai [28] or Wax and Kailath [23] used AIC, MDL or EDC type criteria if the errors are known to be Gaussian. If we know that the errors are Gaussian, we should use that information and that should yield better results. We recommend to use the modified EDC as follows. Suppose

$$I(k, D_N) = -logL_k + D_N\{k(2p-k)+1\}, \qquad (7.1)$$

here

$$L_k = \frac{\prod_{i=k+1}^{p} \hat{\lambda}_i^N}{\left(\frac{1}{p-k}\sum_{i=1}^{p}\hat{\lambda}_i\right)^{N(p-k)}}$$

and D_N satisfies the following conditions:

$$(a)\ \lim_{N\to\infty}\frac{D_N}{N} = 0 \quad (b)\ \lim_{N\to\infty}\frac{D_N}{loglogN} = \infty. \qquad (7.2)$$

Let

$$\hat{q} = argmin I(k, D_N),$$

then $\hat{q}$ is an estimate of q. Zhao, Krishnaiah and Bai [28] proved that $\hat{q}$ is a consistent estimator of q. Note that

$$P[\hat{q} \neq q] = 1 - P[\hat{q} = q],$$

where

$P[\hat{q} = q]$
$= P[I(q, D_N) - I(k, D_N) < 0, \text{ for } k = 0, \ldots, q-1, q+1, \ldots, p-1].$ (7.3)

Since (7.3) depends on the distribution of $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$, using the asymptotic distribution of $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$, we can estimate (7.3) exactly as before by using the re-sampling technique. From a class of D_N, we can estimate (7.3) for each penalty function D_N and choose that D_N which

gives the smallest estimated probability of wrong detection. We use the same model of Section 5, and use the following D_N, defined as $D(1),\ldots,D(12)$ satisfy (7.2) (except D(4) = 1, which gives AIC). They are
$D(1) = N^{\cdot 1}$, $D(2) = N^{\cdot 5}$, $D(3) = N^{\cdot 9}$,
$D(4) = 1$, $D(5) = \frac{1}{4}log(N)$, $D(6) = (log(N))^{\cdot 1}$,
$D(7) = (log(N))^{\cdot 5}$, $D(8) = (log(N))^{\cdot 9}$,
$D(9) = (Nlog(N))^{\cdot 1}$,
$D(10) = (Nlog(N))^{\cdot 5}$,
$D(11) = (Nlog(N))^{\cdot 9}$, $D(12) = \frac{1}{2}log(N)$.

Out of one thousand replications, the percentage of correct estimates and the percentage of wrong estimates are reported in Table 3.

Table 3

$\sigma = .75$

Method	PCE	PWE
New	0.999	0.001
AIC	0.884	0.116
BIC	0.999	0.001
EDC	0.927	0.073

$\sigma = 1.00$

Method	PCE	PWE
New	0.990	0.010
AIC	0.887	0.113
BIC	0.982	0.018
EDC	0.929	0.071

$\sigma = 1.125$

Method	PCE	PWE
New	0.973	0.027
AIC	0.890	0.110
BIC	0.864	0.136
EDC	0.928	0.072

From the results of Table 3, it is very clear that the performances of BIC, EDC and modified EDC improve as the SNR increases. It again verifies the consistency properties of the three methods. On the other hand the inconsistency of the AIC also verifies from the results of Table 3. The performance of BIC is generally better than AIC or EDC if the SNR is moderate (σ = 1.00) or high (σ = .75) but the performance of BIC becomes quite bad compared to AIC or EDC at low SNR (σ = 1.25). The same phenomena were observed in [11] also. Now comparing the modified EDC with the rest it is clear that the modified EDC performs much better than the other known methods at all SNR. It may not be very surprising, because AIC, BIC and EDC use data independent penalty functions. On the other hand the modified EDC use data dependent penalty function. It uses that penalty function, which is in some sense optimal for that given data set within that given class of penalty functions.

8. CONCLUSIONS:

In this paper we consider the direction of arrival model and propose a new method to estimate the number of signals. We do not need any distributional assumptions on the error random variables, except the finiteness of the second order moment. It is well known that if the error are Gaussian, then we can use the AIC, MDL or EDC to estimate the number of signals. We propose a modified criteria, if it is known that the errors are Gaussian. It is observed that our proposed method works better than the usual AIC, MDL or EDC in many situations. It may be mentioned that the proposed method may be difficult for online implementation. More work is needed in that direction.

ACKNOWLEDGMENTS: Most of the work was done when the author was visiting the Department of Statistics, The Pennsylvania State University as a visiting associate professor. The author would like to thank the Department for kind hospitality and also to Professor C.R. Rao for many useful suggestions. The author would like to thank two referees for their constructive comments.

REFERENCES:

[1] Bai, Z.D. and Rao, C.R. (1989), "Spectral analytic methods for the estimation of signals and direction of arrival", *Spectral Analysis in One or Two dimensions*, Eds. Prasad, S. and Kashyap, R.L., Oxford and IBH Publications Co., New Delhi, pp. 493-507.

[2] Bhandari, M and Bansal, N. (1991), "Bayes estimation of the number of signals", *Ann. of Inst. Stat. Math.*, **43**, 2, pp. 227-243.

[3] Bienvenu, G. and Kopp, L. (1980), "Adaptivity of background noise spatial coherence for high resolution passive methods", *Proc. ICASSP 80*, Denver, CO, pp. 307-310.

[4] Brillinger, D.R. (1981), *Time Series Data Analysis and Theory*, Holden-Day Inc., San Francisco, USA.

[5] Bronez, T.P. and Cadzow, J.A. (1983), "An alge-

braic approach to super resolution array processing", *IEEE Trans. Aero. Elec. Syst.*, **19**, No. 1, pp. 123-133.

[6] Johnson, D.H. and Degraff, S.R. (1982), "Improving the resolution of bearing in passive sonar arrays by eigen analysis", *IEEE Trans. Acoust. Speech Signal Processing*, **29**, pp. 401-413.

[7] Kannan, N. (1992), "Estimation of direction of arrival of signals in signal processing", *Ph.D. Thesis*, Department of Statistics, The Pennsylvania State University, University Park, USA.

[8] Kannan, N., Kundu, D. and Mitra, A. (1995), "Estimating DOA of signals", *Signal Processing and Communications*, Eds. Makur, A. and Reddy, V.U., Tata McGraw Hill, New Delhi, pp. 63-68.

[9] Kaveh, M., Wang, H. and Hung, H. (1987), "On the theoretical performance of a class of estimators of the number of narrow band sources", *IEEE Trans. ASSP*, Vol. 35, pp. 1350-1352.

[10] Kumaresan, R. and Tufts, D. (1983), "Estimation of arrival of multiple plane waves", *IEEE Trans. Aero. Elec. Syst.*, **19**, pp. 134-139.

[11] Kundu, D. (1995), "Small sample properties of some parametric and non parametric methods for detection of signals by Monte Carlo simulations", *Proc. IEE Vision, Image and Signal Processing*, **142**, No. 3, pp. 181-186.

[12] Kundu, D. (1996), "A modified MUSIC algorithm to estimate the DOA of signals", *Signal Processing*, **48**, No. 1, pp. 85-90.

[13] Kundu, D. (1998), "Estimating the number of sinusoids and its performance analysis", *Jour. of Stat. Comp. and Simul.*, Vol. 60, pp. 347-362.

[14] Kundu, D. (1999), "Estimating direction of arrival of signal; some asymptotic results", *Sankhya, A*, Vol. 61, No. 1, pp. 120-138.

[15] Owsley, N.L. (1977), "Spectral signal set extraction", *Aspects of signal processing*, Part II, Ed. Eacconi, G., pp. 469-475.

[16] Paulraj, A., Otteresten, B., Roy, R., Swindlehurst, A., Xu, G and Kailath, T. (1993), "Subspace methods for direction of arrival estimation", *Handbook of Statistics*, Eds. Bose, N.K. and Rao, C.R., Vol. 10, , Chpater 16, 693-739, Elsevier Science Publishers.

[17] Pillai, S.U. (1989), *Array Signal Processing*, Springer Verlag, New York.

[18] Press, W.H., Teukolsky, S.A., Vellerling, W.T. and Flannery, B.P. (1992), *Numerical Recipes in FORTRAN, The Art of Scientific Computing*, 2nd. Edition, Cambridge University Press.

[19] Reddy, V.U. and Biradar, L.S. (1993), "SVD-Based information theoretic criteria for detection of the number of signals of the damped/undamped sinusoids and their performance analysis", *IEEE Trans. Signal Processing*, **41**, 9, pp. 2872-2881.

[20] Roy, R., Paulraj, A. and Kailath, T. (1986), "ESPRIT - A Subspace Rotation Approach to Estimation of Parameters of Cisoids in Noise", *IEEE Trans. on ASSP*. Vol. 34, pp. 1340-1342.

[21] Schmidt, R.O. (1981), "A signal subspace approach to multiple emitter location and spectral estimation", *Ph.D. Thesis*, Stanford University.

[22] Wang, H. and Kaveh, M. (1986), "On the performance of signal subspace processing, Part I, narrow band systems", *IEEE Trans. Acoust. Speech Signal Processing*, **34**, pp. 1201-1209.

[23] Wax, M. and Kailath, T. (1985), "Detection of signals by information theoretic criteria", *IEEE Trans. Acoust. Speech Signal Process.*, **33**, No. 2, pp. 387-392.

[24] Wax, M., Schmidt, R.O. and Kailath, T. (1983), "Eigen-structure method for retrieving the poles from the natural response", *Proc. 2nd. CDC*, TX, pp. 1343-1344.

[25] Wax, M., Shan, T.J. and Kailath, T. (1984), "Spatio temporal spectral analysis by e-eigen-structure methods", *IEEE Trans. Acoust. Speech Signal Process.*, **32**, No. 4, pp. 817-827.

[26] Wilkinson, W.H. (1965), *Algebraic Eigenvalue Problems*, Oxford University Press.

[27] Yin, Y.Q. and Krishnaiah, P.R. (1987), "On some non-parametric methods for detection of the number of signals", *IEEE Trans. Acoust. Speech Signal*

Processing, **35**, pp. 1533-1538.

[28] Zhao, L.C., Krishnaiah, P.R. and Bai, Z.D. (1986), "On the detection of number of signals in the presence of white noise", *Journal of Multivariate Analysis*, **20**, 1, pp. 1-25.